**Задание для первого студента.** Распарсить сайт из списка согласно номеру задания и создать web-интерфейс для отображения полученных данных. Краулер должен считывать новостную ленту с первой страницы сайта. Периодичность повторения устанавливается пользователем. Данные заполняются в БД MongoDB. Обязательные поля для текста новости:

- Название новости

- Дата новости

- Ссылка на новость

- Текст новости

- Ссылка на видео (если есть)

- Количество просмотров новости (если есть)

- Количество комментариев новости (если есть)

При очередном проходе краулера для существующих в БД новостей (определяется по Название новости, Дата новости, Ссылка на новость) поля количества просмотров и

Вариант 9.  https://vlg-media.ru/

**Задание для второго студента.**

Создать программный модуль для анализа новостей из БД. Выделить с помощью Томита-парсера упоминание в тексте значимых персон Волгоградской области и достопримечательностей. Зафиксировать в БД соответствующие предложения (с указанием на новость-источник) для дальнейшего анализа тональности. Вывести в общем web-интерфейсе.

Персоны https://global-volgograd.ru/person

Достопримечательности https://avolgograd.com/sights?obl=vgg

**Задание для третьего студента.**

Создать программный модуль для выявления тональности высказываний по отношению к предложениям, с упоминанием персон Волгоградской области и достопримечательностей. Использовать методы машинного обучения на размеченном корпусе. Вывести в общем web-интерфейсе.

Создать программный модуль анализа модели word2vec на всем объеме новостных статей из БД. Для персон Волгоградской области и достопримечательностей определить контекстные синонимы и слова, с которыми они упоминались в тексте.